Krieger, N., & Davey Smith, G. (2017). Response: FACEing reality: productive tensions between our epidemiological questions, methods, and mission. *International Journal of Epidemiology*, *45*(6), 1852-1865. https://doi.org/10.1093/ije/dyw330

Title: **FACEing reality: productive tensions between our epidemiological questions, methods, and mission**

Authors

Nancy Krieger, PhD, Professor of Social Epidemiology, American Cancer Society Clinical Research Professor, Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health

George Davey Smith, MD, DSc, Professor of Clinical Epidemiology, MRC Integrative Epidemiology Unit at the University of Bristol

Corresponding author

Nancy Krieger, PhD
Professor of Social Epidemiology
American Cancer Society Clinical Research Professor
Department of Social and Behavioral Sciences, Kresge 717
Harvard T.H. Chan School of Public Health
677 Huntington Avenue
Boston, MA 02115 (USA)
office: 617-432-1571; fax: 617-432-3123
email: nkrieger@hsph.harvard.edu

Word count 6565 (text)

Tables (textboxes) 3

Submitted to *Int J Epidemiol* on August 29, 2016
Revised and resubmitted: October 16, 2016

Funding. NK was supported, in part, by her American Cancer Society Clinical Research Professor Award. GDS works within a Unit supported by the MRC (MC-UU-12013/1) and the University of Bristol.

We appreciate the commentaries provoked by our paper [1]. Our intent was to flesh out unacknowledged assumptions and problematic framings and practices pertaining to "causal inference" in our field of epidemiology, and we did so with full recognition that our debates are not unique and have counterparts in virtually all disciplines involved in testing empirical claims about the world and universe in which we humans find ourselves [2-4]. The commentaries indicate we have accomplished our purpose, whereby the articles, considered together, productively point to important tensions between our epidemiologic questions and methods, as they relate to our mission of generating knowledge vital for improving population health and promoting health equity.

In our reply, we address where we agree and disagree with each commentary. Rather than rehash arguments provided in our original paper, we elaborate only on the substantive disagreements that we consider important to the discussion underway.

Our starting point, one well-recognized both by practicing scientists and by those who study scientific disciplines, is that use of valid methods and their development is essential for scientific advancement, even as no science, including epidemiology, is defined solely by its methods [2-4]. Improved methods are essential to expand the types of data we can obtain and how we analyze them, increasing the odds for generating valid inferences. They can, as several of the commentaries stress, also help clarify which empirical questions we currently have the material capacity to ask and answer (as related to technologies of discovery and informatics), and those which we cannot – thereby motivating new work to make asking the currently unanswerable questions more tractable, as well as provide better answers to questions already thought to be answered.

We can never lose sight, however, that the questions we even think to pose, or believe we can ask, depend, first and foremost, on the theories, concepts, and material evidence that inform each and every scientific effort to explain the material workings of our world [4-6]. In our field, involving dynamic populations of people in dynamic societies and ecosystems, methodical triangulation of diverse types of

evidence from diverse types of study designs and involving diverse population is essential [7,8], as is recognition that any "causal effect" detected in a given empirical study is the effect estimated in that study – it is not "the" (as in: "universal") "causal effect." Indeed, as Sander Greenland and Jamie Robins observed back in 1988 [9, p. 1195]:

> "…the dependency of epidemiologic measures on cofactor distributions points out the need to avoid considering such measures as biological constants. Rather, epidemiologic measures are characteristics of particular populations under particular conditions …"

Testing hypotheses under diverse conditions (e.g., study designs that use different methods with uncorrelated biases and that conducted in diverse populations with different prevalence of exposures and confounders), i.e., systematic triangulation of approaches to posing scientific questions – and listening to the material world "talk back," not solely statistical manipulation of data from any given study – is the surest route to challenge our explanations and gaining new robust knowledge about how this world works to produce the patterns of health, disease, and well-being we seek to analyze and alter. It is for this reason that we maintain that a critical orientation to evidence, explanation, and methods is essential, as provided by the incisive and flexible framework of "Inference to the Best Explanation" [10].

Or, as the Professor of Social Biology and statistician, experimental embryologist, and evolutionary biologist Lancelot Hogben stated in the conclusion to his 1933 opus *Nature and Nurture*, which robustly challenged the era's dominant eugenics orientation to the analysis of population health [11, p. 121]:

> " The application of statistical technique in the study of human inheritance is beset with pitfalls. On the one hand the experimental difficulties of the subject-matter necessitate recourse to mathematical refinements which can be dispensed with in animal breeding. On the other there is the danger of concealing assumptions which have no factual basis behind an impressive façade of flawless algebra. The student may well recall the words of Wilhelm Ostwald:
>
> > Among scientific articles there are to be found not a few wherein the logic and mathematics are faultless but which for all that are worthless, because the assumptions and hypotheses upon which the faultless logic and mathematics rest do not correspond to actuality."

And with this caveat in mind, we turn to our responses to the individual commentaries [12-16].

Response to commentaries

1) Blakely, Lynch, and Bentley: "DAGs and the restricted potential outcomes approach are tools, not theories of causation" [12]

We agree: that DAGs can be very useful tools, that tools are not theories – and that critical deep substantive knowledge is required to use these tools well. As Blakely et al. note, use of DAGs can be "meaningless and perhaps even misleading without a well formed theory of why one node causes another" [12]. We also agree, per their Figure 1, that science necessarily involves epistemology – defined, by the Oxford English Dictionary as: "the theory of knowledge and understanding, esp. with regards to its methods, validity, and scope, and the distinction between justified belief and opinion" [17]. Equally necessary, however, is engaging with ontology: "the science or study of being" [17], referring to the understandings of the phenomena and processes of our world, including their causal relationships and the apt contrasts that can make them visible [4,5,10]. If our paper succeeds in cautioning against, in Blakely et al.'s words, a "blinkered approach" to using DAGs while clarifying where and how DAGs can be used well in "situations where it is apposite to apply" [12], we are satisfied.

2) VanderWeele: "On Causes, Causal Inference, and Potential Outcomes" [13]

We appreciate five key points raised by VanderWeele's thoughtful commentary. These pertain to the commentary's:

(1) distinction between what may be theorized to be a "cause" and what may, more narrowly, comprise a "causal effect estimand" (point #1);

(2) recognition that epidemiology tackles many problems that involve "composite exposures," thereby making it "difficult to speak of *the* causal effect estimand," a problem further compounded by it being difficult to uphold the stringent assumption of "no unmeasured confounding" (point #3), let alone no misclassification or uncertainty in measurement! – which are critical issues we address

further on, including in **Textboxes 1, 2 and 3.**

(3) acknowledgement of the value, in the case of composite exposures, of considering the impact of multiple types of interventions and also the value of recognizing when it is appropriate to "abandon any attempt at a precise interpretation of a quantitative estimand and simply speak of evidence for general causation, for something to be a cause" (point #3);

(4) realization that there are "important questions that are not amenable to the potential outcomes framework" (such as health impacts of "social movements, societal trends such as more married women going to work, and even war"), whereas the latter is more apt for "narrower policy evaluations and decisions" (point #4); and

(5) acceptance that that "scientific reasoning is much broader than causal inference," that "inference to the best explanation is important in causal inference and diverse types of evidence can and should be used," and that there can be substantial disagreement among scientists as to what constitutes the "best explanation," a matter that cannot be resolved solely by use of counterfactual reasoning, however powerful such reasoning can be (point #6).

The larger message, which we support, is that it is essential to recognize the limits of any particular set of methods and the questions that can be asked using them, without dismissing as unimportant questions that cannot (currently) be asked using these methods.

We disagree, however, with some key aspects of how the commentary frames scientific research and knowledge. Specifically, in point #2, the commentary presents science as being an activity in which: (1) "investigators collect data – encoded as strings of numbers," (2) "a data analyst performs various computations on these strings," after which "the investigators may produce written or verbal sentences concerning causal relationships," a practice which requires that the investigators use "a formal framework for translation" to discern "the relationship between these strings of numbers and the resulting sentences." In point #3, it privileges, for causal analysis, the physical sciences and "laws of

nature" against the study of "complex systems" in the "social and biomedical sciences."

Yet, "data" are never simply a "given," something that are simply "collected" [4-6]. Instead, and as we recognize VanderWeele and other epidemiologists deeply appreciate, substantial causal theorizing in relation to critical substantive knowledge is required for conceptualizing the variables and hypotheses in the first place, not to mention choice of study design and comparison groups, let alone operationalizing and actually measuring the conceptual variables of interest [4-6]. Scientific inquiry does not start with "data": even so-called hypothesis free approaches (such as genome wide association studies [18] or phenome wide association studies [19] are predicated on an understanding of the structuring of genotype and phenotype.

Additionally, not only are there fierce debates over what are "laws of nature" and whether they can be legitimately conceptualized as the epitome of "cause" [4,5,20-22], but physical sciences are not immune to engaging with complexity, context, and historical contingency: consider only the evolution of such complex non-living physical systems and phenomena such as galaxies, stars, black holes, planets and planetary systems. Life, of course, introduces the additional complexity of dynamic living beings interactively engaging (instinctively and/or consciously) with other living beings, both in the same and different species, while simultaneously actively constructing their niches as they grow, develop, eat, relate, sometimes reproduce, and ultimately die [6,23]. In fields where dynamic and context-dependent phenomena comprise the domain of study, such as epidemiology or ecology, few exposures could ever be expected to yield precise "causal estimands" that are indifferent to the actual contexts in which people and other organisms live. Of course there are exceptions: no human fetus will survive if is missing one of its chromosomes #1 (regardless of environmental context)[24], and no humans will survive if they breathe carbon monoxide at concentrations of greater than 200 ppm for prolonged intervals (regardless of their genome)[25]. Most of what epidemiology studies, however, occupies the far more messy turf of non-deterministic exposures and outcomes that are context-dependent (to a greater or lesser degree,

depending on the phenomenon under study), and our methods must engage accordingly.

It is therefore surprising that VanderWeele writes (point #1) that "For a hypothetical intervention it [i.e. the potential outcomes approach] defines the causal effect for the individual." Such a stance is at odds with the essence of epidemiology: that it is a population science [6,26,27]. We contend that the notion of individual causal effects is fundamentally non-epidemiological: for the vast majority of epidemiological endpoints the notion of an individual causal effect is simply incoherent [27-29]. Granted, there are some deterministic exposures that invariably lead to the same consequence in all individuals, as per the two examples provided above (regarding a missing chromosome #1 or breathing too much carbon monoxide), as well as more complex scenarios, e.g., the onset of severe neurocognitive anomalies persons with Huntington's disease for those carrying a sufficient number of expansions, as compared to persons who do not carry these expansions (given survival until age of onset) [30]. That deterministic models can be rewritten as probabilistic ones does not detract from the conceptual distinction between population-level and individual-level causal effects, a distinction that underpins epidemiological science [26,27,31].

VanderWeele makes strong claims for the potential outcomes approach, stating that: (a) it has led to "[truly] major advances over what was available previously" (point #2) and (b) "What little progress has been made ... has in fact come out of extensions of the potential outcomes or causal diagram framework" (point #3) [13]. However the assumptions made in the application of these approaches are – as VanderWeele partly acknowledges – nothing short of heroic. The notion of "no unmeasured confounding" (point #3)[13] not only assumes perfect measurement even for those confounders that have been measured, but also is an ideal that rarely, if ever, is reached in any epidemiologic analyses of real-world (as opposed to simulated) data. Underscoring the underappreciated nature of this problem, a recent cogent review on causal inference for research using high dimensional biological data (which minimally is what comprehensive epidemiological data on

exposures and outcomes are!) perceptively observed that the assumption that "variables are measured without measurement error" is a "a subtle assumption that is required … [in causal analysis methods] often not realized by practitioners who apply these techniques" [32]. Being clear about the limitations and assumptions of methods is crucial to preventing invalid inference, as well as to identifying issues requiring more methodological development.

Finally, the commentary errs in suggesting that: (a) we think the problem is the dismissal of "race" as a cause (e.g., by those who argue that "causes" in a counterfactual framework must be "manipulable"), and that (b) this problem can be righted by acknowledging " 'race is a cause' in the sense that whatever might be meant by race, or however one might define or conceptualize race, that conception would entail also certain features such as skin color …" (point #7) [13].

As we explicitly stated, and as the commentary quotes, "the relevant counterfactual pertains to racism, not 'race.'" We will not here restate the myriad reasons why it is invalid to conceptualize "race" as inherent trait, let alone reduce "race" to a matter of "skin color" (think only of the popular, albeit complex, saying in Brazil that "money whitens" [33,34]). Nor will we restate the quotes provided (in Textbox 4 of our original paper) of researchers asserting that "race" cannot be a "cause" in a counterfactual framework because it is not "manipulable," a stance that necessarily and erroneously assumes "race" is an *a priori* intrinsic trait. Our point, worth restating, concerns the necessity of epidemiologic research seriously grappling with analyzing how racism harms health, including by constructing the notion of "race" as an inherent biological variable, which apparently then becomes (at least to some) immune to counterfactual interrogation.

Here we note that we are heartened that VanderWeele supports research using counterfactual reasoning to examine how racism harms health, as evidenced by the study he mentions in which one of us (NK) collaborated. Indeed, clarifying how "race" was conceptualized, the text in the first paragraph of this paper explicitly stated: "Here, we adopt a counterfactual causal inference framework to investigate

determinants of black–white disparities as proposed by VanderWeele and Robinson, framed by an understanding that inequitable race relations, not "race" per se, are the cause of racial/ethnic health inequities, that is, unjust, unfair, and preventable social inequalities in health ” [35, p. 83]. We strongly encourage more epidemiologists to take on these types of questions, using counterfactual-based methodologies where apt, but not letting methods alone drive the research questions they pose.

3) Weed: “Causal inference in epidemiology: potential outcomes, pluralism and peer review **[14]**

If Weed deems scientific critique as an exercise of “academic tower” egos only (to “gain attention”; to incite “intellectual competitions”), and thinks graduate instruction should be geared to simplified rules, and that scientific reviews need only take into account a study’s methods to assess the merits of its causal claims, that is his prerogative. We disagree.

Additionally, with regard to Weed’s concern that we are raising a false alarm over a non-existent danger, see our response to Robins and Weissman [15], below, and also Textboxes 1-3.

4) Robins JM, Weissman MB. “Counterfactual causation and streetlamps: what is to be done?” **[15]**

Offering a refreshingly different take on the value of productive scientific debate and exchange, we welcome Robins and Weissman’s vigorous engagement with the issues we epidemiologists and other scientists face in the difficult work of framing and testing causal claims. That they think the stakes are high is evidenced by the second clause of their title: we venture to guess that many readers of *Int J Epidemiol* might not be aware that “What is To Be Done?” is intended to evoke *What is To Be Done? Burning Questions of our Movement*, written in 1902 by Vladimir Ilyich Ulyanov (later to adopt the pseudonym Lenin), in which he argued that only a revolutionary party, not solely working class organizations (such as trade unions), could lead the kind of revolution needed to overthrow the Czar and

the rest of the Russian state and replace it with, in the terminology of the day, a Marxist Social-Democratic state [36]. To some, Ulyanov's tract precipitated the split of the Russian Social Democratic Labour Party (RSDLP) into the Bolshevik and Menshevik factions a year later. While we are curious whom in the current debate Robins and Weissman cast as the Bolsheviks and as the Mensheviks (or, for that matter, the Okrana [the state secret police], or the vacillating Georgi Plekhanov, elder statesman of the RSDLP in dismay as his authority washed away), we are happy to let those particular sleeping dogs lie (!).

To start, we are encouraged to see clear acknowledgment of the "streetlamp" problem, i.e., that "recent causal methods illuminate some potential treatments much more directly than they illuminate others" [15]. Types of problems outside of their arc of light include those that, in the parlance of political and policy scientists, are "wicked" problems, i.e., involve complex systems whose very properties can be changed by virtue of the interventions introduced – and issues of health inequities certainly belong to this category of problem [37-39]. Restricting a focus to only "well-behaved" problems, however, is not a solution, especially since, as Robins and Weissman acknowledge, turning on or off the "lamps" of better research methods by itself does not determine the location of the "keys" that can help generate solutions.

Better instead to take on the challenge of making less tractable problems more tractable, at least to the point of being able to come to some general statements about "causation," per VanderWeele's formulation [13], even if generating a precise "causal estimand" is not possible. Here, the reflective warnings of Robert May, whose training is in physics, zoology, and ecology (and who has served as President of the Royal Society in the UK, and also Chief Scientific Advisor to the UK Government and Head of the UK Office of Science and Technology [40]) are instructive [41, p. 196]:

> "In the USA, the National Science Foundation (NSF) went through a rather irritating hiccup from the mid 1970s to the mid 1980s when a bunch of people (who didn't understand physics but thought that ecology ought to be more like what they thought physics was) decided that the NSF ecology programme really ought to be doing reductionist things, and reductionism got confused with doing

> manipulative experiments. The outcome of this was an interesting phase: if you look at the manipulative experiments in ecological studies published over this time, 75% were of a spatial scale of less than 10 metres and 95% of them were on a time scale of five years or less, the time scale of a PhD thesis. It is not at all clear that the most important questions in understanding community structure and response to disturbance necessarily happen on a scale of less than a metre or over a timescale of three years or less."

The truth of these words is ever more apparent in this era of global climate change, and it is no accident that May has served on the UK Government's Climate Change Committee.

We find it interesting that Robins and Weissman treat "Inference to the Best Explanation" [10] as the "standard approach to science." Would that it were so. As we recount, it is in fact a contested approach [2-4,10] – and, as we noted, one barely mentioned in the epidemiological literature [1]. Moreover, regarding triangulation, their analogy of meaninglessly "triangulating" across the different "truth" meanings of the word "gay," used differently in different contexts to mean entirely different things, is erroneous. In the case of empirical research in which evidence is being triangulated, the studies are presumably seeking to generate what ought to be roughly similar – though not precisely the same - causal estimands, if it is indeed true that the underlying theorized causal relationship actually occurs in the shared biophysical world we inhabit. This is a completely different scenario from that of the same "word" (i.e., same set of letters) having different meanings, precisely because of how language evolves and changes (see, for example, Raymond Williams' classic text *Keywords* for an historically grounded study of etymology in cultural context [42]).

The example of HDL cholesterol that Robins and Weissman discuss illustrates several issues we have with their claim – echoed in the commentaries by VanderWeele [13] and Daniel et al [16] – that the methods advocated produce estimates of causal effects that are not only "unavailable by other techniques" but, beyond this, can provide "more or less reliable, relatively assumption-free estimates of relatively well-defined causal effects that are unavailable by other techniques" [15]. The saga of HDL cholesterol and cardiovascular disease, as we explain in **Textbox 1**, suggests otherwise – and instead underscores our view that major advances in identification of causal processes that allow for individual

and public health improvement depend considerably more on advances in understanding the material basis of human biology than in how we draw diagrams of such processes [43].

From this standpoint, we also find it instructive that Robins and Weissman find our choice of empirical examples to be puzzling. All of the examples share a common point: the need for deep substantive knowledge about the theorized causal processes. Several also highlight surprising new scientific observations to serve as a reminder of how we must always temper our claims about "causal estimands" because we cannot know if all relevant causal relationships and confounders have been taken into account. From this standpoint, we thoroughly endorse the statement Robins and Weissman quote that is attributed to Keynes: "When my information changes, I alter my conclusions. What do you do, sir?" [15]. (Would only that such an attitude towards evidence were evident in current policy and political discourse …).

The larger point is that data are never simply "strings of numbers" (as per VanderWeele's formulation [13]). Instead, application of methods to data, including counterfactual causal methods, can only contribute to generating robust knowledge if there is clear theorizing about the nature (often dynamic) of the variables involved and their possible causal relations. Cogent DAGs could have been drawn for etiologically incorrect explanations of pellagra, and the elimination and addition of variables from and to the picture came about because of deep biological knowledge combined with causal reasoning (including but not limited to counterfactual reasoning). Of course one would wish this to be the standard approach. But consider only the legions of studies and teaching examples, including those using DAGs, that assume "race" is "simply" an intrinsic "biological" variable and that never consider asking how racism simultaneously harms health and delimits not only the "racial" categories employed but who is put in them, and by whom [6,44-46]. A DAG won't make a researcher see the gaps: theory and deep substantive knowledge of the complexities of the phenomena under study are vital, as is knowledge about the historical controversies concerning their study [4-6].

Finally, we welcome Robins and Weissman's closing examples pertaining to studying such questions as: "How will our lives differ if we build suburbs versus dense walkable cities?"; and also: what is the impact of greenhouse gases on climate change? [15] Their examples include the kind of triangulation of evidence from diverse sources that we were advocating, and recognize that any estimates of health impact of complex exposures and policies will have uncertainty. That is the world we live in – and the world in which we are trying to improve population health and promote health equity. As responsible scientists, we not only have to be clear about uncertainty and the limitations of our knowledge, but also be clear about trying to get the best evidence we can, especially for understudied (and often controversial) topics that require more light – or even light of different wavelengths – than the illumination provided by the proverbial streetlamps. By the same token, we have to expose the restrictive assumptions (along with typically poor evidence) that inevitably feed into dominant accounts that justify inequity. We make no presumption that social movements and "social variables" necessarily point in the direction of health equity: certainly the Nazis had a mass base and plenty of some types of "social capital."

These tensions between our field's questions, methods, and mission are not new, and nor do we claim new insights about their existence and resolution. What we do offer is a reminder that these tensions exist, a reminder that needs to be repeated in each and every generation, and we are glad to contribute our part. Or, as Edgar Sydenstricker concluded, in his 1933 landmark book *Health and Environment* [47, p. 210]:

> "What is needed is more knowledge, dispassionately collected and scientifically analyzed with a wholesome respect for the complexities of human societies and of the individuals who compose it, to form a sound basis for the conscious control of our destinies."

5) Daniel, De Stavola, Vansteelandt: "The formal approach to quantitative causal inference in epidemiology: misguided or misrepresented?" **[16]**

There has long been a sense that an appreciation for the material and complex realities of the

issues we analyze in epidemiology may be obscured by an exclusive focus on the methods of how we conduct our analyses [6]. We see this disconnect in the commentary by Daniel and colleagues [16]. Approaching 40 years ago, in the early days of this journal, Edmond Murphy in 1978 wrote eloquently and explicitly about the disconnect between biological and statistical understandings, stating that "multivariate analysis (which in certain quarters is being substituted for scientific perception), can spread its soporific effect" and also that (with respect to some analyses) "I am driven to believe that however excellent the prediction, the formula, from an aetiological and ontological standpoint, provides no insights whatsoever" [48]. Two other leading epidemiologists expressed similar views around the same time: in 1980 Reuel Stallones lamented that recent work in epidemiology demonstrated a "continuing concern for methods, and especially the dissection of risk assessment, that would do credit to a Talmudic scholar and that threatens at times to bury all that is good and beautiful in epidemiology under an avalanche of mathematical trivia and neologisms" [49], a view Jerry Morris shared [50].

We quote these earlier contributions (which bear re-reading today) as they illustrate the sense that method was coming to dominate over matter in the epidemiological enterprise. Our analysis of historical examples, far from giving license to "conclude that science does not need formal deductive theory at all" – as asserted by Daniel et al [16] -- instead shows that science needs not only theories pertaining to methods but theories pertaining to substance, as well as an openness to learning from history (if only not to repeat prior errors) [2-6].

First, however, we are glad to see that Daniel et al. do not hold that "hypothetical interventions must be currently humanly feasible" or that the randomized clinical trial is always "the best choice of study design for causal inference" [16]. One corollary of the first statement is their view that the causal effects of inherited genes can be studied, since a particular mutation "could instead hypothetically not have been inherited," even if inducing such a mutation (or preventing it) is "currently unfeasible to implement" [16].

By contrast, we find Daniel et al.'s discussion about health inequities involving "sex, race, and genes" to be confused, and this problem is compounded by their dismissal of concerns raised as "something of a storm in a teacup" [16]. Such a stance implies unfamiliarity with the literature on health inequities, let alone the toll they take on people's lives [44-46,51,52]. This approach to the subject matter stands in stark contrast to their stated credo that a "Formal Approach to quantitative Causal inference in Epidemiology" (FACE) necessarily recognizes "the central role played by subject-matter knowledge" [16].

Yet, pointing to serious limitations in "subject-matter knowledge" [16], Daniel et al. describe "race" solely in genetic terms, and as examples of racial groups refer to "Caucasian" and "Afrocaribbean" [16]. Never mind that the term "Caucasian" is a discredited pseudo-scientific term, invented by Blumenbach in the late 1700s and premised on the idea that humanity originated in the mountainous area of the Caucuses, near Mount Ararat (the Biblical resting place of Noah's ark), and that the other "races" of humankind descended (or, in Blumenbach's terminology, "degenerated") from this original stock of humanity [53-56]. Such a thesis not only flies in the face of current evidence strongly supporting the out-of-Africa origins of *Homo sapiens* but also obscures how terms like "Caucasian" put a seemingly scientific gloss on the unwarranted assumption that "race" is an *a priori* intrinsic biological property of individuals [54-60]. It also deflects attention from who gets to count, according to whom, as being either "Caucasian" (or "white") or "Afrocarribean" (or "black"): both Europe and the Caribbean have a muddy and bloody history over who has been counted, by whom, as belonging to which categories, and approaches to classification and self-identification can differ by country, by island, by historical period, and, among emigrants, by the country in which they newly reside [33,34,57-62]. For an article that repeatedly stresses the importance of intellectual precision, the laxity of its approach to "race" and racism is striking indeed. Related, to suggest that the issue of "exchangeability" merely entails considering whether an individual is raised by a family of the same or different "racial" group

entirely misses the point of analyzing structural racism as a determinant of population health: what would happen if people lived in a society premised on racial justice rather than one which in which racial discrimination flourishes?

The suggestion that what primarily matters is "the *perception* of race/sex" is not defensible. Such a statement ignores, for instance, the impact of the intergenerational accumulation – within both families and the neighborhoods and societies in which they reside – of privilege vs. adversity. At issue are entangled and entrenched racial/ethnic inequities in wealth, neighborhood living conditions, and safety, co-occurring with gender inequities in education, occupation, and income, along with interpersonal and structural violence shaped by class, race, and gender relations [44-46,51,52]. These are the circumstances into which infants are born and which they inherit, long before anyone has perceived them.

It is additionally a misreading of what we wrote to suggest we were implying that "race" can and should be construed as a "cause" once the objection that it is not "manipulable" is removed. In fact, we said exactly the opposite. Our argument is that "race" should NOT be considered a "cause" – not because of the arguments about it being "non-modifiable," but because: (a) "race" is not a valid biological category [57-60], and (b) the societal relations and practices that constitute structural racism along with its interpersonal expressions are the causal phenomena of interest [44-46]. Of concern is how biologically embodying racism in its many manifestations (which, yes, can be and are carefully parsed in the literature) has profound effects on people's health [44-46,51,52]. There is an extensive literature on analyzing racism and gender inequality as determinants of health and health inequities [6,44-46,51,52], which provides the "subject-matter knowledge" as to who and what may cause racial/ethnic and gender health inequities. Such knowledge needs to be assimilated, both substantively and theoretically, if an informed view of how they should be analyzed is to be produced.

Our alleged misconceptions about the roles of DAGs in causal inference are additionally

puzzling. To reiterate, we think DAGs can often be useful, and (as Daniel et al. appear to agree with) that serious substantive knowledge and causal theorizing is required for a DAG to be valuable, rather than potentially misleading. However DAGs should not lead the search for causes, because even if the tendency is resisted, there will be an inevitable move to studying what can be reduced to a tractable DAG (just as when you have the proverbial hammer, every problem can look like a nail). This concern, of course, applies to all approaches aimed at strengthening causal inference, from negative controls to Mendelian randomization and other instrumental variables methods.

We further note that in our experience -- and contrary to what Daniel et al. assert [16] -- many of the DAGs presented in the literature or in the classroom do not adhere to the high standard of: (a) reflecting deep subject area knowledge; (b) being presented in all plausible forms (including taking into account the potential problem of equivalence, whereby different causal DAGs can be consistent with closely similar or identical correlation structures [63]); (c) fully acknowledging the possibility of rapid feedback networks that challenge meaningful acyclicity; and (d) incorporating all possible measurement imprecision and unmeasured confounders, whilst being tractable. In **Textboxes 2 and 3**, we present two examples – respectively pertaining to C-reactive protein and cardiovascular disease and to education and diabetes, among many that could have been selected – in which the use of DAGs does not adhere to these standards and which also make no attempt at triangulation or inference to the best explanation (IBE).

As Elwert has observed, "one obvious challenge of working with DAGs is that the true causal DAG is often not known," and "[t]his is a problem because identification always hinges on the validity of the causal model" [64]. No matter how nicely drawn the DAG (as now facilitated by the tidy program DAGitty [65] – see **Textbox 3**), if its underlying theoretical premises are unsound, so too will be its causal estimands. If our paper encourages more critical thought as to what goes into DAGs (or is left out), so much the better.

Indeed, the DAGs that Daniel et al. present in their commentary are illustrative of the problems at issue. First, for the pellagra example, their Figure 1 misses the point that the initial DAGs that would have been drawn would never have included all the elements incorporated into their Figure; one cannot do Whig science and impose current knowledge on the past. In the pre-Goldberg era, the two sorts of DAGs that would have been drawn (had anyone drawn DAGs back then; Sewall Wright was only just beginning to innovate in this direction, with his work on path analysis in the 1920s [63,66], including his recognition of the need to include anchors that could allow understanding of direction of cause [43]), would have focused, respectively, on: (a) contaminated food, and (b) infectious etiology, and both would have been missing the key unobserved variable that was in fact the true causal agent: vitamin deficiency, as set into motion by the political economy of sharecropping and institutionalization of impoverished orphans, persons with disabilities, and prisoners. Putting the variables we mentioned into one DAG, and perhaps allowing DAGitty to suggest a "minimally sufficient adjustment set" [65], is unlikely to have yielded the correct answer in this case.

The analysis of the DAG for the "birthweight paradox" (Figure 2) is equally problematic. The 2014 commentary by VanderWeele [67] that Daniel et al. mention, and which we affirmatively cited in favor of the position we were arguing, was published only after DAGs had been employed to infer that the "paradox" could be explained away, methodologically, as opposed to engaged with, biologically [68]. As we discussed in our paper, the value of VanderWeele's commentary is that it recognizes that similar DAGs could encode different causal scenarios (i.e., the underappreciated "equivalence" problem [63]). This is a major advance over initial views that DAGs help demonstrate that the "paradox" is in effect a methodological artifact, whereby use of the DAGs was stated to show that "this apparent paradox can be characterized as selection bias due to stratification on a variable (birthweight) that is affected by the exposure of interest (smoking) and those share common causes with the outcome (infant mortality)." [68]. From the telling of the tale by Daniel et al., however, DAGs uncomplicatedly have pointed to an

adequate solution all along, which demonstrably is not the case. Ignoring the history of a problem is not a path to better knowledge.

Concluding thoughts: theory & rendering the invisible visible

Our paper and the various commentaries do share a common theme: the need for rigorous thoughtful science that can address the myriad questions of our times – and also alert us to future challenges. Three commentaries, moreover, come close to surmising why we may have written our paper as we have, at this present time:

1) Commenting on the current enthusiasm for the potential outcomes approach for analyzing structural relationships between variables, Blakely et al. warn that "when teaching epidemiology it would be unfortunate to solely rely on DAGs and counterfactual approaches," since this would lead to an "unfortunate and unhelpful restriction of epidemiology's scope, but more importantly, would limit understanding of how the health of a population is shaped" [12].

2) VanderWeele likewise observed that the "popularity of the potential outcomes approach within epidemiology" may have resulted in questions "which are not amenable to a potential outcomes analysis … perhaps receiving less attention," a problem which in turn means it is important to ensure "that, within teaching, and in the published literature, more examples of broader questions concerning systems and movements are discussed" [13].

3) Robins and Weissman similarly note "[S]ince the theory and application of formal counterfactual causal methods is undergoing rapid and novel development, it is natural that many such papers are being published," such that "current methodological research literature naturally appears highly skewed towards papers on counterfactual causality" [15].

Add these statements to all three commentaries' recognition that many important social problems with major implications for population health cannot readily be analyzed by these increasingly popular

methods, and the problem should be clear. As other scientific fields recognize that share these same tensions [2-5, 41], we can confine our field to the small plot of terrain near the proverbial streetlamp – or we can recognize that a commitment to discovery and explanation of determinants of population health and health inequities additionally requires us to expand the scope and venture further. No dog relies on sight alone.

In 1935, Major Greenwood, the first-ever Professor of Epidemiology and Vital Statistics at the London School of Hygiene and Tropical Medicine [69] published his classic book *Epidemics and Crowd Diseases: An Introduction to the Study of Epidemiology* [70]. It notably encompassed not only diseases of infectious etiology, the dominant focus of late 19th and early 20th c CE epidemiology, but also touched upon diseases involving nutritional, occupational, and psychological causes. Knowledge regarding the latter outcomes, however, was much more scarce, because, in Greenwood's words, "the bacteriological school had become psychologically omnipotent" [70, p. 60]. As he acutely observed, there was not "any logical reason why identification of *contagia viva* should lead us to discard general epidemiological principles … but … such was the *practical effect* of the discovery" [70, p. 60].

The warning is clear. Much as we welcome the strengthening of epidemiologic knowledge through application of counterfactual causal reasoning and methods, both of which are the products of deep scientific insight and hard work, we know – from what we see around us now, from our knowledge of our discipline's history, and from our broader study of the history and philosophy of science – that it would be counterproductive to limit epidemiology solely to use of this framework and the questions it can answer with precision.

Acknowledgement: NK dedicates this paper to the memory of Ruth Hubbard (1924-2016), her mentor since college, who taught her to think critically about science, to do science critically, to challenge whenever biology is invoked as an excuse for injustice, and to think expansively about the people's health in relation to history and social justice.

Textbox 1. The saga of HDL cholesterol & heart disease: what led to improved causal inference from observational data?

To briefly recap the long-running story: whether HDL (high density lipoprotein, popularly known as "good cholesterol") protects against coronary heart disease (CHD), and thus treatments that raise HDL would reduce CHD risk, has been investigated for decades. By the late 20th c CE, a general consensus emerged in favor of this hypothesis, leading to the launching of large-scale randomised clinical trials (RCTs) of HDL cholesterol raising, to the cost of many hundreds of millions of dollars [71,72].

In 1991, two decades before these RCTs were initiated, one of us (GDS) co-authored a paper which argued that it was not possible, using observational data, to generate separate reliable estimates of any putative effects of HDL and triglycerides, given issues involving analysis of imprecisely measured highly correlated exposures, in particular triglyceride levels [73]. Nevertheless, observational studies of HDL which statistically adjusted for triglyceride continued to appear, and in 2009 an influential international collaborative effort reported that the apparently protective HDL effect was robust to covariate (including triglyceride) adjustment, whereas the apparently detrimental triglyceride effect was not [74]. The authors concluded that the "current findings suggest that therapy directed at HDL-C as well as non-HDL-C may generate substantial additional benefit" [74], an unambiguous prediction as to what RCTs of HDL cholesterol raising would show. A paper by pioneers of the causal inference field used HDL and heart disease as a hypothetical example [75], and various attempts at improved inference through propensity scores (e.g. [76]) were carried out, but the status of HDL cholesterol as a protective factor remained the majority view.

Yet, only a few years later, these conclusions based on observational data were challenged by new data. What changed was not new ways of analyzing conventional observational data – or representing the problem graphically. Instead, the advance was the introduction of an approach predicated on recent advances in molecular genetics – Mendelian randomization (MR), which takes advantage of "the random assortment of genes from parents to offspring that occurs during gamete formation and conception" [77]. Using genetic variants as instrumental variable was advocated as a potentially more robust approach to causal inference in the kind of situation exemplified by HDL cholesterol, triglyceride and CHD [78]. At this time common genetic variants with robust association with HDL cholesterol were not available.

However, as the relevant genetic variants were identified (largely through genome wide association studies), MR studies suggested - against the substantial and largely consistent naïve observational data analyses - that HDL levels *per se* were not protective against CHD [79-81]. This finding was in agreement with an ever-increasing list of large-scale RCTs of several different pharmacological approaches to raising HDL cholesterol [71,72,82]. This is not a unique situation: several other cases have appeared in which observational epidemiological and other data suggested there would be a protective effect of pharmacologically manipulating a biomarker, but MR studies and RCTs agreed in their null findings (to the cost of yet more hundreds of millions of dollars)[83]. A rapidly growing array of MR studies are appearing [84], including ones in which MR analyses have been reported before the RCTs testing the same causal process [85], and it is doubtful that the HDL debacle would be repeated in an age when such MR studies can be done.

Thus, contrary to the commentaries' claims that "extensions of the potential outcomes or causal diagram framework" are key for "what little progress has been made" in tough epidemiological problems [13], yielding robust inference "unavailable by other techniques" [15], critically improved understanding has come from advances in material understandings of how the world is, and ways of harnessing the power of perturbations to the material world, not idealized diagrams. The text by Lenin that Robins and Weissman should perhaps have instead mentioned in this regard is not "What is to be done" [36] but rather "Materialism and empirio-criticism" [86], where the advantages of a materialist understanding over the then prevalent idealism is (in somewhat blunt terms) advocated.

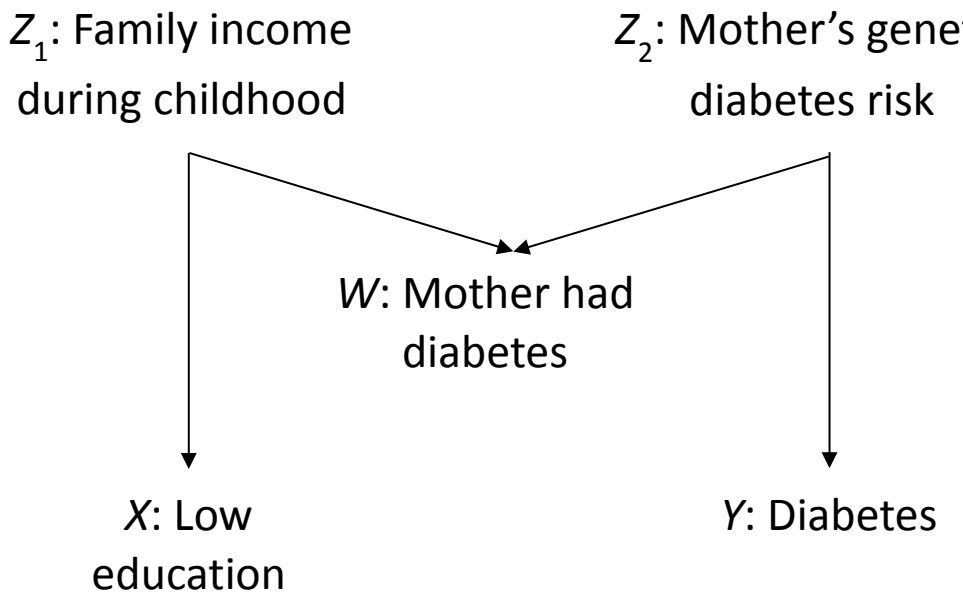
TEXT BOX 2. *Meaningless "causal estimands": the case of C-reactive protein and cardiovascular disease*

A study published in *Epidemiology*, in some ways the house-journal of the FACE movement, and acknowledging the help of Tyler VanderWeele, presents a DAG-encoded analysis of which factors mediate the effect of higher body mass index (BMI) on coronary heart disease (CHD) [87]. The analysis identifies the acute phase reactant C-Reactive Protein (CRP) as a mediator. However, "deep subject matter" knowledge of the field would strongly question the role of CRP as a mediator, since to be a mediator a factor must causally influence the outcome. Yet, a substantial body of evidence makes clear that CRP is not causal with respect to CHD [88-90]. Epidemiologic analyses that have wrongly inferred CRP is a causal factor have been plagued by the unsurprising set of problems that are always threats to valid causal inference in epidemiology [91], including: (a) measurement error; (b), the strong and essentially statistically entirely intractable effect of reverse causality (in this case, atherosclerosis, an inflammatory process, and CRP levels co-evolve over time, with atherosclerosis strongly influencing CHD risk); and (c) confounding, from a myriad of factors.

There is, however, no discussion of these problems in the paper at all. Instead, per the quantitative emphasis of "the Formal Approach to quantitative Causal inference in Epidemiology" (FACE) [16], the article avers that CRP accounts for 6-8% of the effect of BMI on CHD [87]. Granted, this is a nicely quantitative causal estimand. It is, however, meaningless, if CRP cannot, biologically, be a mediator. Failure to detect this fundamental causal problem was likely abetted by the lack of use of either triangulation or inference to the best explanation (IBE), contrary to the suggestion, by several of the commentators on our paper [15,16], that both triangulation and IBE are merely the common-sense that epidemiologists already utilize near universally. Despite these problems, a commentary on the paper celebrated its use of DAGs and FACE with the title "Beyond vague causal effect estimation of obesity on health outcomes" [92] – but is it really an advance to have precise but meaningless estimands? As with the HDL cholesterol example discussed in **Textbox 1**, other approaches to mediation analysis that use causal anchors rather than statistical manipulation can produce more reliable evidence on mediation [93,94], as demonstrated by an analysis of the same question as addressed in the paper under discussion [95].

Textbox 3. Diabetes, education, and DAGitty: algorithms are not a substitute for subject-matter expertise.

In a chapter promoting the use of causal diagrams in social epidemiology Maria Glymour presented a DAG representing the causal null for the hypothesis that education has no effect on diabetes, [96], as follows:



She imagines the situation where we have measures of mother's diabetes status, but do not have measures of family income when the individuals were growing up, or if the individual's mother had elevated genetic risk for diabetes. She asks the question whether we should adjust are analyses from mother's diabetes status, following the "graphical criteria" for whether a factor is a confounder (e.g. see introduction to these in Shrier and Platt [97]). The conclusion is that adjusting for mother's diabetes would introduce a spurious statistical association between low education and diabetes, as this would be conditioning on a collider, and therefore "under the graphical criteria, one should not include mother's diabetes status as a covariate".

However an equally plausible hypothesis is that maternal diabetes – which would be reflected in hyperglycaemia and other within-pregnancy factors that could influence fetal brain development – could influence cognitive development and this attainment of low educational status of her offspring (a suggestion for which there is a body of evidence [98]). This would be classical confounding, and in this situation not adjusting for maternal diabetes would leave potentially substantial confounding, as clearly maternal diabetes would be a confounder for the association between low education of the offspring and the offspring risk of diabetes (as maternal diabetes will influence offspring diabetes through both germline genetic transmission and potentially through intrauterine influences). Glymour does not mention this alternative possibility.

How then, does the "central role played by subject-matter knowledge" that Daniel et al mention [16] play out in this situation, and influence the potentially vast number of DAGS that could be drawn for this question? Should a form of sensitivity analysis be performed which include carrying out an analyses for every possible DAG? And for the DAGs that haven't been thought of? And for the DAGs containing measures for which you have no data? We recognise that these are central issues for epidemiology, and that the statement that everything should be done in the spirit of sensitivity analysis is a reasonable one (although such sensitivity analyses are dependent on subject-area knowledge and will not be implementable for the last two of the above options). However what will happen in practice?

We consider it likely that the automatic use of DAGs will not enhance "the central role played by subject matter knowledge" [16], nor is the notion that DAGs currently serve as a help-mate to triangulation how we read the current literature. Instead, DAGs are often constructed that conveniently require precisely the data the authors have to hand (and not data they cannot access), that are vitiated by measurement error and obvious unmeasured

confounders, that do not feed their analyses into any attempt at triangulation or inference to the best explanation, and that produce the same tired examples of "this week's random medical news" as has damaged the reputation of epidemiology for years [91]. As one case in point, a cross-sectional analysis of periodontal infection and glucose intolerance was centered around alternative DAGs [99], and produced the same likely residually confounded associations as would be expected from such data [91,100,101], yet concluded that if replicated the "public health implications would be substantial", with the presence of and discussion of the DAGs somehow shoring up very weak evidence [99].

Our skepticism in regard to Daniel et al's notion that FACE approaches are "always guided by what [Krieger and Davey Smith] call IBE (inference the best explanation)" [16] is increased by observing how convenient and useful programs for constructing DAGs, such as DAGitty [65,102] are used. The program has an automated approach to finding so-called "minimally sufficient adjustment sets" [65, 103] and many papers now contain a version of the statement "we used the software DAGitty to find a minimally sufficient adjustment set"(e.g. [104-107]). We invite readers to examine these and many other such papers (that can be located through following up the 200+ papers that cite DAGitty publications) to determine their own view as to the application of subject matter knowledge and triangulation to causal inference, and evaluate whether Daniel et al's statements in this regard match reality. Selection of the covariates will of course be driven by data availability, together with measurement characteristics of the variables that are available, and sample size (that will influence the apparent conditional independencies that contribute to such selection). Unsurprisingly, implementation of one of these DAGitty selected models suggests that CRP is a mediator between BMI and cognitive function [106] (just as with the suggestion CRP mediates between BMI and CHD, discussed in **Textbox 2**), with no acknowledgement of the evidence not supporting CRP causally influencing cognitive function [108].

DAGitty also offers automation with respect to both: (1) mediation analysis (with all the problems inherent in this, see **Textbox 2**) and (2) the selection of instrumental variables from the data provided, i.e., selecting a variable with an open path to the exposure of interest X, and for which all paths between it and the outcome are closed by {X} [65]. It will also locate instruments that only meet these criteria when conditioned on a set of covariates. The use of instruments that are data-derived, rather than from subject-specific knowledge, is likely to lead to highly misleading findings, given the impact of measurement characteristics on such selection. The fact that such selection of instruments has been previously advocated [110] does not render it reliable. The semi-automation of these analyses will not, in our view, improve the practice of causal inference in epidemiology.

REFERENCES

1. Krieger N, Davey Smith G. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *Int J Epidemiol* (in press).

2. Illari PM, Russo F, Williamson J (eds). *Causality in the Sciences*. New York: Oxford University Press, 2011.

3. Beebee H, Hitchcock C, Menzies P. *The Oxford Handbook of Causation*. Oxford: Oxford University Press, 2009.

4. Bird A, Ladyman J (eds). *Arguing about Science*. New York: Routledge, 2012.

5. Ziman JM. *Real Science: What it is, and What it Means*. Cambridge, UK: Cambridge University Press, 2000.

6. Krieger N. *Epidemiology and the People's Health: Theory and Context*. New York: Oxford University Press, 2011.

7. Lawlor DA, Tilling K, Davey Smith G. Triangulation and aetiological epidemiology. *Int J Epidemiol* (in press).

8. Richmond RC, Al-Amin A, Smith Davey G, Relton CL. Approaches for drawing causal inferences from epidemiological birth cohorts: a review. *Early Hum Development* 2014; 90:769-780.

9. Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol* 1988; 128:1185-1197.

10. Lipton P. *Inference to the Best Explanation*. 2nd ed. London: Routledge, 2004.

11. Hogben L. *Nature and Nurture*. New York: W.W. Norton & Co., Inc., 1933.

12. Blakely T, Lynch J, Bentley R. DAGs and the restricted potential outcomes approach are tools,

not theories of causation. *Int J Epidemiol* (in press).

13. VanderWeele TJ. On causes, causal inference, and potential outcomes. *Int J Epidemiol* (in press).

14. Weed DL. Causal inference in epidemiology: potential outcomes, pluralism and peer review. *Int J Epidemiol* (in press).

15. Robins JM, Weissman MB. Counterfactual causation and streetlamps: what is to be done? *Int J Epidemiol* (in press).

16. Daniel RM, De Stavola BL, Vansteelandt S. The formal approach to quantitative causal inference in epidemiology: misguided or misrepresented? *Int J Epidemiol* (in press).

17. *OED Online*. Oxford University Press, June 2016. Web. 16 October 2016.

18. McCarthy MI, Abescasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics* 2008; 9:356-369.

19. Millard LAC, Davies NM, Timpson NJT, Tilling K, Flach PA, Davey Smith G. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. *Scientific Reports* 2015; 5:16645. doi:10.1038/srep16645

20. Carroll JW. Laws of Nature. In: Zalta EN (ed). *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition). http://plato.stanford.edu/archives/spr2016/entries/laws-of-nature/ . 16 October 2016.

21. Lange M. Laws of nature. In: Skalr L (ed). *Physical Theory: Method and Interpretation*. Published to Oxford Scholarship Online: October 2014. DOI: 10.1093/acprof:oso/9780195145649.001.0001

22. Bird A. *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press, 2007.

23. Gilbert SF, Epel D. *Ecological Developmental Biology: The Environmental Regulation of*

Development, Health, and Evolution. Sunderland, MA: Sinauer Assoc., 2015.

24. van den Berg MMJ, van Maarle MC, van Wely M, Goddinj M. Genetics of early miscarriage. *Biochimica et Biophysica Acta* 2012; 1822:1951-1959.

25. Roderique JD, Josef BS, Feldman MJ, Spiess BD. A modern literature review of carbon monoxide poisoning theories, therapies, and potential targets for therapy advancement. *Toxicology* 2015; 334:45-58.

26. Krieger N. Who and what is a "population?' Historical debates, current controversies, and implications for understanding "population health" and rectifying health inequities. *Milbank Quarterly* 2012; 90:634-681.

27. Davey Smith G. Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. *Int J Epidemiol* 2011; 40:3:537-562.

28. Davey Smith G, Relton Cl, Breannan P. Chance, choice, and cause in cancer aetiology: individual and population perspectives. *Int J Epidemiol* (in press).

29. Newman TB. Etiology of ventricular septal defects: an epidemiological approach. *Pediatrics* 1985; 76:741–749.

30. Reiner A, Dragatsis I, Dietrich P. Genetics and neuropathology of Huntington's disease. *Int Rev Neurobiol* 2011; 98:325-372.

31. Rose G. Sick Individuals and Sick Populations. *Int J Epidemiol* 1985; 14:32-38.

32. Lagani V, Triantafillou S, Ball G, Tegnér J, Tsamardinos I. Probabilistic Computational Causal Discovery for Systems Biology. In: Geris L, Gomez-Cabrero D (Eds). *Uncertainty in Biology Volume 17 of the series. Studies in Mechanobiology, Tissue Engineering and Biomaterials* 2016; 33-73.

33. Nobles M. History counts: a comparative analysis of racial/color categorization in US and Brazilian censuses. *Am J Public Health* 2000; 90:1738–1745.

34. Telles E, Paschel T. Who is black, white, or mixed race? How skin color, status, and nation shape racial classification in Latin America. *Am J Sociol* 2014; 120:864-907.

35. Valeri L, Chen JT, Garcia-Albeniz X, Krieger N, VanderWeele TJ, Coull BA. The role of stage at diagnosis in colorectal cancer black–white survival disparities: a counterfactual causal inference approach. *Cancer Epidemiol Biomarkers Prev* 2016; 25:83-89.

36. Lenin VI. What is To Be Done? Burning Questions of Our Movement. 1902. In: *Lenin: Selected Works in Three Volumes*. Moscow: Progress Publishers, 1977; Volume 1:92-241.

37. Wistow J, Blackman T, Byrne D, Wistow G. *Studying Health Inequalities: An Applied Approach*. Bristol, UK: Policy Press, 2015.

38. Balint PJ, Stewart RE, Desai A, Walters LC. *The Challenge of Wicked Problems: Managing Uncertainty and Conflict*. Washington, DC: Island Press, 2011.

39. Dorfman L, Krasnow ID. Public health and media advocacy. *Annu Rev Public Health* 2014; 35:293-306.

40. Robert May, University of Oxford, Department of Zoology. http://www.zoo.ox.ac.uk/people/view/may_r.htm ; 16 October 2016.

41. May R. Levels of organization in ecological systems. In: Bock GR, Goode JA. The Limits of Reductionism in Biology. (Novartis Foundation Symposium 213, May 1998). Chichester, UK: John Wiley & Sons, 1998; pp. 193-202.

42. Williams R. *Keywords: A Vocabulary of Culture and Society*. Revised ed. New York: Oxford University Press, 1983.

43. Davey Smith G. The Wright stuff: genes in the interrogation of correlation and causation. *Eur J Personality* 2012; 26:395-397.

44. Krieger N. Discrimination and health inequities. In: Berkman LF, Kawachi I, Glymour MM. *Social Epidemiology*. 2nd ed. New York: Oxford University Press, 2014; pp. 63-125.

45. Paradies Y, Ben J, Denson N, Elias A, Priest N, Pieterse A, et al. Racism as a determinant of health: a systematic review and met-analysis. *PLoS ONE* 2015; 10(9):e0138511.

46. Metzl J, Roberts D. Structural competency meets structural racism: race, politics, and the structure of medical knowledge. *The Virtual Mentor: VM* 2014; 16:674-690.

47. Sydenstricker E. *Health and Environment*. New York: McGraw Hill, 1933.

48. Murphy EA. Epidemiological strategies and genetic factors. *Int J Epidemiol* 1978; 7:7-14.

49. Stallones RA. To advance epidemiology. *Annu Rev Public Health* 1980*;* 1**:**69*–82.*

50. Morris JN. Modern epidemiology? *J Epidemiol Community Health* 1988*;* 42**:**100*.*

51. Hofrichter R, Bhatia R (eds). *Tackling Health Inequities Through Public Health Practice: Theory To Action*. 2nd ed. New York: Oxford University Press, 2010.

52. Hunt K, Annandale E (eds). *Gender and Health: Major Themes in Health and Social Welfare*. New York: Routledge, 2012.

53. Blumenbach F. On the Natural Variety of Mankind (1775). In: Blumenbach JF. *The Anthropological Treatises of Johann Friedrich Blumenbach*. London: Published for the Anthropological Society, by Longman, Green, Longman, Roberts & Green, 1865; pp. 65-145.

54. Augstein HF. From the land of the Bible to the Caucuses and beyond: the shifting ideas of the geographical origins of humankind. In: Ernst W, Harris B (eds). *Race, Science, and Medicine, 1700-1960.* London: Routledge, 1999; pp. 58-79.

55. Krieger N. Stormy weather: “race,” gene expression, and the science of health disparities. *Am J Public Health* 2005; 95:2155-2160.

56. Painter NI. *The History of White People*. New York: W.W. Norton & Co., 2010.

57. Marks J. The nature/culture of genetic facts. *Annu Rev Anthropol* 2013; 42:247-267.

58. Wailoo K, Nelson A, Lee C. *Genetics and the Unsettled Past: The Collision of DNA, Race, and History*. New Brunswick, NJ; Rutgers, 2012.

59. Koenig B, Lee S, Richardson S (eds). *Revisiting Race in a Genomic Age*. New Brunswick, NJ: Rutgers University Press, 2008.

60. Yudell M. Race *Unmasked: Biology and Race in the Twentieth Century*. New York: Columbia University Press, 2014.

61. Waters MC, Ueda R (ed), with Helen B. Marrow. *The New Americans: A Guide to Immigration since 1965*. Cambridge, MA: Harvard University Press, 2007.

62. Anderson W. Racial conceptions in the Global South. *Isis* 2014; 105:782-792.

63. Shipley B. *Cause and Correlation in Biology: A User’s Guide to Path Analysis, Structural Equations, and Causal Inference*. Cambridge, UK: Cambridge University Press, 2000.

64. Elwert F. Graphical causal models. In: Morgan SL (ed). *Handbook of Causal Analysis for Social Research*. Dordrecht: Springer Science + Business, 2013; pp. 245-273.

65. Textor J. Drawing and analyzing causal DAGs with DAGitty. (User manual for DAGitty version 2.3, released 19 August 2015). http://dagitty.net/ ; 16 October 2016.

66. Crow JF. *Sewall Wright, 1889-1988: A Biographical Memoir*. Washington, DC: National Academy of Sciences, 1994.

67. VanderWeele TJ. Commentary: resolution of the birthweight paradox: competing explanations and analytical insights. *Int J Epidemiol* 2014; 43:1368-1373.

68. Hernández-Díaz S, Schisterman EF, Hérnan MA. The birth-weight paradox uncovered? *Am J Epidemiol* 2006; 164:1115-1120.

69. Hogben L. Major Greenwood. 1880-1949. Obituary Notices of Fellows of the Royal Society. *Royal Soc* 1950; 7:139-154.

70. Greenwood M. *Epidemics and Crowd Diseases: An Introduction to the Study of Epidemiology*. London: Williams & Norgate, 1935.

71. Lloyd-Jones DM. Niacin and HDL cholesterol – time to face facts. *NEJM* 2014; 371:271-273.

72. Kingwell BA, Chapman MJ, Kontush A, Miller NE. HDL-targeted therapies: progress, failures and future. *Nature reviews Drug discovery* 2014; 13:445-464.

73. Phillips AN, Davey Smith G. How independent are "independent" effects? Relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol* 1991; 44:1223-1231.

74. The Emerging Risk Factors Collaboration. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* 2009; 302:1993-2000.

75. Spirtes P, Scheines R. Causal inference of ambiguous manipulations. *Philosophy of Science* 2004; 71:833-845.

76. Seo SM, Choo EH, Koh YS, Park MW, Shin DI, Choi YS, Park HJ, Kim DB, Her SH, Lee JM, Park CS. High-density lipoprotein cholesterol as a predictor of clinical outcomes in patients achieving low-density lipoprotein cholesterol targets with statins after percutaneous coronary intervention. *Heart* 2011; 10:hrt-2011.

77. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; 32:1-22.

78. Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004; 33:30-42.

79. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, Dale CE, Padmanabhan S, Finan C, Swerdlow DI, Tragante V. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart Journal* 2014; Jan 28:eht571.

80. Haase CL, Tybjærg-Hansen A, Ali Qayyum A, Schou J, Nordestgaard BG, Frikke-Schmidt R. LCAT, HDL cholesterol and ischemic cardiovascular disease: a Mendelian randomization study of HDL cholesterol in 54,500 individuals. *J Clin Endocrinol Metab* 2011; 97:E248-56.

81. Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, Hindy G, Hólm H, Ding EL, Johnson T, Schunkert H. Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 2012; 380:572-80.

82. Rader DJ. New therapeutic approaches to the treatment of Dyslipidemia. *Cell Metabolism* 2016; 23:405-412.

83. Talmud PJ, Holmes MV. Deciphering the Causal Role of sPLA2s and Lp-PLA2 in Coronary Heart Disease. *Arteriosclerosis Thrombosis Vascular Biol* 2015; 35: 2281-2289.

84. Boef AG, Dekkers OM, le Cessie S. Mendelian randomization studies: a review of the approaches used and the quality of reporting. *Int J Epidemiol* 2015; 44:496-511.

85. Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2× 2 factorial Mendelian randomization study. *J*

Am Coll Cardiol 2015; 65:1552-1561.

86. Lenin V. *Materialism and Empirio-Criticism*. New York: International Publishers, 1927. .

87. Lu Y, Hajifathalian K, Rimm EB, Ezzati M, Danaei G. Mediators of the effect of body mass index on coronary heart disease: decomposing direct and indirect effects. *Epidemiology* 2015; 26:153-162.

88. C Reactive Protein Coronary Heart Disease Genetics Collaboration. Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. *BMJ* 2011; 342:d548.

89. Elliott P, Chambers JC, Zhang W, Clarke R, Hopewell JC, Peden JF, Erdmann J, Braund P, Engert JC, Bennett D, Coin L. Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA* 2009; 302:37-48.

90. Zacho J, Tybjærg-Hansen A, Jensen JS, Grande P, Sillesen H, Nordestgaard BG. Genetically elevated C-reactive protein and ischemic vascular disease. *NEJM* 2008; 359:1897-908.

91. Davey Smith G, Ebrahim S. Epidemiology—is it time to call it a day? *Int J Epidemiol* 2001; 30:1-1.

92. Chiolero A. Commentary: beyond vague causal effect estimation of obesity on health outcomes. *Epidemiology* 2015; 26:163-164.

93. Richmond RC, Hemani G, Tilling K, Davey Smith G, Relton CL. Challenges and novel approaches for investigating molecular mediation. *Human Molecular Genetics* 2016; Jul 20:ddw197.

94. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol* 2012; 41: 161-176.

95. Varbo A, Benn M, Davey Smith G, Timpson NJ, Tybjærg-Hansen A, Nordestgaard BG. Remnant

cholesterol, low-density lipoprotein cholesterol, and blood pressure as mediators from obesity to ischemic heart disease. *Circulation Res* 2015; 116:665-673.

96. Glymour MM. Using causal diagrams to understand common problems in social epidemiology. In: Oakes JM, Kaufman JS (eds). *Methods in Social Epidemiology*. San Francisco: Josey-Bass, 2006; pp. 393-428.

97. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Method* 2008; 8:70. doi: 10.1186/1471-2288-8-70.

98. Adane AA, Mishra GD, Tooth LR. Diabetes in Pregnancy and Childhood Cognitive Development: A Systematic Review. *Pediatrics* 2016; 137(5):e20154234.

99. Arora N, Papapanou PN, Rosenbaum M, Jacobs DR, Desvarieux M, Demmer RT. Periodontal infection, impaired fasting glucose and impaired glucose tolerance: results from the Continuous National Health and Nutrition Examination Survey 2009–2010. *J Clin Periodontol* 2014; 41:643-652.

100. Shungin D, Cornelis MC, Divaris K, Holtfreter B, Shaffer JR, Yu YH, Barros SP, Beck JD, Biffar R, Boerwinkle EA, Crout RJ. Using genetics to test the causal relationship of total adiposity and periodontitis: Mendelian randomization analyses in the Gene-Lifestyle Interactions and Dental Endpoints (GLIDE) Consortium. *Int J Epidemiol* 2015; 44:638-650.

101. Borgnakke WS. Modifiable Risk Factors for Periodontitis and Diabetes. *Current Oral Health Rep* 2016; 3:254–269.

102. Textor J, Hardt J, Knüppel S. DAGitty: A Graphical Tool for Analyzing Causal Diagrams. *Epidemiology* 2011; 5:745.

103. Textor J, van der Zander B, Gilthorpe MS et al. Robust causal inference using Directed Acyclic

Graphs: the R package “dagitty”. *Int J Epidemiol* (in press).

104. Kowall B, Lehnich AT, Strucksberg KH, Führer D, Erbel R, Jankovic N, Moebus S, Jöckel KH, Stang A. Associations among sleep disturbances, nocturnal sleep duration, daytime napping, and incident prediabetes and type 2 diabetes: the Heinz Nixdorf Recall Study. *Sleep Med* 2016; 21:35-41.

105. Röhrig N, Strobl R, Müller M, Perz S, Kääb S, Martens E, Peters A, Linkohr B, Grill E. Directed acyclic graphs helped to identify confounding in the association of disability and electrocardiographic findings: results from the KORA-Age study. *J Clin Epidemiol* 2014; 67:199-206.

106. Gunathilake R, Oldmeadow C, McEvoy , Inder KJ, Schofield PW, Nair BR, Attia J. The association between obesity and cognitive function in older persons: how much is mediated by inflammation, fasting plasma glucose, and hypertriglyceridemia? *J Gerontol A Biol Sci Med Sci* 2016. pii: glw070.

107. Hinkle SN, Sharma AJ, Kim SY, Schieve LA. Maternal prepregnancy weight status and associations with children's development and disabilities at kindergarten. *Int J Obes* (Lond). 2013; 37:1344-1351.

108. Mooijaart SP, Sattar N, Trompet S, Polisecki E, de Craen AJM, Schaefer EJ, Jahn SE, van Himbergen T, Welsh P, Ford I, Stott DJ, Westendorp RGJ, on behalf of The PROSPER Study Group. C-Reactive Protein and genetic variants and cognitive decline in old age: The PROSPER Study. *PLoS ONE* 2011; 6(9): e23890. doi:10.1371/journal.pone.0023890

109. Marioni RE, Deaery IJ, Murray GD, Lowe DGO, Ransson SB, Strachan MWJ, Luciano M, Houlihan LM, Gow AJ, Harris SE, Stewart MC, Rumley A, Fowkes FGR, Price JF. Genetic variants associated with altered plasma levels of C-Reactive Protein are not associated with late-life cognitive ability

in four Scottish samples. *Behavior Genetics* 2010; 40:3-11.

110. Tchetgen EJ, Lin SH. Robust estimation of pure/natural direct effects with mediator measurement error. Harvard University Biostatistics Working Paper Series, Paper 152; 2012. http://biostats.bepress.com/harvardbiostat/paper152 . 16 October 2016.